Phonological motivation for the acquisition of onomatopoeia:

An analysis of early words

Catherine E. Laing

Centre for Language and Communication Research,

Cardiff University

laingc@cardiff.ac.uk

ORCiD ID: 0000-0001-8022-2655

Twitter: @cathelaing24

This research was funded by the Economic and Social Research Council (Grant No. ES/J500215/1).

Abstract

Onomatopoeia are disproportionately high in number in infants' early words compared to adult language. Studies of infant language perception have proposed an iconic advantage for onomatopoeia, which may make them easier for infants to learn (Imai & Kita, 2014; Laing, 2017). This study analyses infants' early word production to show a phonological motivation for onomatopoeia in early acquisition. Cross-linguistic evidence from 16 infants demonstrates how these forms fit within a phonologically-systematic developing lexicon. We observe a predominance of consonant harmony and open CV syllables in infants' early words – structures that are typical of onomatopoeia across languages. Infants' acquisition of onomatopoeia is shown to be driven by a preference for structures that are easy to plan and produce. These data present an original perspective on onomatopoeia in early development, highlighting the role of production in language acquisition in general, and onomatopoeic words in particular.

Keywords: onomatopoeia, phonological development, lexical development, early production, iconicity.

Phonological motivation for the acquisition of onomatopoeia: An analysis of early words

Onomatopoeia – that is, words that represent sounds from the environment, such as animal and engine noises, and words such as *bang*, *boom* and *yum yum* – are especially prominent in infants' earliest words. In Menn and Vihman's (2011) appendix of 48 infants' first five words (across 10 different languages), 20% were onomatopoeic. Kern (2010) reported similar findings: words classed on the French adaptation of the MacArthur CDI (Kern & Gayraud, 2010) as 'sound effects and animal sounds' were the most common in infants' productive repertoires up to 20 months old, contributing to over a third of the entire output until age 16 months. Tardif and colleagues' (2008) cross-linguistic study also shows a striking presence of onomatopoeia in early words; on average the category of 'sound effects' was second only to 'people' (e.g. *mummy*, *daddy*) in 970 infants' most common words across American English (29.5% of all words), Cantonese (40.6%) and Putonghua (8.7%). Nevertheless, onomatopoeia are often omitted from analyses of infants' lexical development (e.g. Behrens, 2006; Fikkert & Levelt, 2008); as Vihman (2014) concludes, "the question of *just how* these less than fully arbitrary transitional forms support early word learning has *seldom been addressed*" (p.161, italics added). In relation to the prominence of onomatopoeia in early production, they have received little attention in infant language research.

There is compelling evidence to show an advantage for iconic forms, including onomatopoeia, in early perception. Perry, Perlman and Lupyan (2015) considered age of acquisition of all words on the English and Spanish Communicative Development Inventories (CDI, Fenson et al., 1994) to find that words acquired earliest tend to be more iconic – that is, more closely related to their meaning (e.g. *woof woof*, *mushy* and *stomp* are more iconic than *better*, *empty* and *smell*; see Winter, Perlman, Perry, & Lupyan, 2017, for iconicity ratings). This result was consistent when onomatopoeia were removed from the data. Laing (2017) supports these findings with evidence from early perception, showing that ten-month-olds are better able to map onomatopoeic words (*woof woof*, *quack quack*) to their referents than non-onomatopoeic words (*dog*, *duck*), so long as those words were familiar from the input.

Werner and Kaplan (1963) suggest a semantic advantage for onomatopoeia, as they present a lexical representation, such as *woof*, that can be more easily related to the form in question (here 'dog') than its arbitrary referent. This does not necessarily depend on any specific iconic attributes of the form, but instead draws upon the more relatable correspondence between an object – usually an animal or a vehicle – and a sound that could be made by that object. But the fact that onomatopoeia represent sounds may be important. Dingemanse (2012) discusses a 'hierarchy' of iconic forms (termed ideophones in his paper), highlighting how sound-meaning correspondences are the most common across languages: "if a language has ideophones at all it will have at least ideophones for sound (i.e. onomatopoeia)" (p.663). He proposes that the mapping between sound and speech is the most obvious in human language, presenting a simple and salient correspondence between a sound in the environment and the sound of a human voice.

While evidence is still only piecemeal, infants' production of onomatopoeia has nevertheless captured researchers' attention. In a study analysing syllabification in Finnish infants' early speech, Kunnari (2002) noted that onomatopoeia were produced more accurately than non-onomatopoeic forms. She posits that this may be due to their "easily mastered articulatory shape" (p.133), which could lead infants to prioritise their acquisition over more complex forms. In another case study, Laing (2014) shows how onomatopoeia provide an alternative option to more challenging target words in the early output of one infant, Annalena (Elsen, 1991). Comparing Annalena's acquisition of onomatopoeia (e.g. *meow*) with their corresponding 'conventional' words[1] (e.g. *cat*), onomatopoeia tended to have simpler segmental properties, which matched the infant's phonetic and phonological capacities in the early stages of word production. Annalena's early words were produced with relative accuracy, and were similar to those of canonical babbling, with CV syllables and reduplication. Onomatopoeia were shown to be a more suitable match to her early production capacity than many non-onomatopoeic words. Over time, Annalena moved towards the adult conventional form, but only when she had sufficient phonological capacity to do so. Based on this study of one infant, it seems that onomatopoeia may be preferentially acquired in early development because they have simple phonological structures that are typical of infants' earliest-acquired words.

Stoel-Gammon and Cooper (1984) discuss the phonological and lexical development of three infants, one of whom, Will, has a strong tendency to use onomatopoeic forms: twenty-one of Will's first 50 words are onomatopoeic, compared with four and zero of the other two infants' words. The authors describe two main production patterns in Will's early output: reduplication, and the use of onomatopoeia to refer to various non-onomatopoeic targets (e.g. *pop-pop* 'fire', *hohoho* 'Santa'). Stoel-Gammon and Cooper (1984) refer to these as "selection patterns" (p.263), as Will remains faithful to the target form by 'selecting' words in early production that matched

[1] As one reviewer noted, the use of the term 'conventional' here might imply that onomatopoeia are unconventional, which is not the case. In this study we take the word 'conventional' to refer to the more 'adult-like' target form that is conventionally associated with the onomatopoeic word.

his phonological capacity. However, the authors do not consider that these two patterns might be related: they note that 17 of Will's first 50 words are reduplicated, in addition to 16 reduplicated onomatopoeic words. In total, 33 of his first 50 words are reduplicated; Will may have acquired onomatopoeia due to their easily-producible forms that match the well-rehearsed structures produced in his output, and not simply because he prefers onomatopoeia. This selective approach to early lexical acquisition is noted in a number of studies. Drawing on observations from Stoel-Gammon (2011), Aoyama and Davis (2017, p.1085) conclude that "lexical development [is] influenced by productive phonology at early stages of…language development". This is supported by numerous findings in the literature. McCune and Vihman (2001) show that a high proportion of early words match consonants produced most frequently in babble, as infants draw from their own phonological repertoire when acquiring their first lexicon. Kim and Davis (2015) show this to be true into later development, as children aged 1;0 to 3;0 consistently matched their word productions with their phonological repertoires; as phonological capacity developed, children began to produce more variable phonological forms. Stoel-Gammon (2011) describes the transition from babble to simple early word structures (e.g. [dada] in babble eventually becoming the meaningful word *daddy* in early speech) as being "a matter of adding meaning to sound" (p.8): early words match the structures of babble, and over time phonological production become more complex as the infant gains production practice and linguistic knowledge.

Work by Vihman and colleagues (Vihman, 2014; Vihman & Croft, 2007; Vihman & Keren-Portnoy, 2013) shows how infants draw upon systematic phonological patterns, or 'prosodic structures', in early language development. First noted by Waterson (1971), accounts demonstrating infants' implementation of systematic phonological structures have contributed a fascinating set of analyses to the phonological development literature. Waterson drew upon

Firth's (1948) 'prosodic analysis' to consider her son's word production, focussing on the whole word as the unit for analysis. This contrasts with the typical segment-by-segment comparison between child and adult form, as seen, for example, in Smith's (1973) account of his son's development. Waterson (1971) noted that, phonologically-speaking, her son's forms were far removed from the adult target, to fit a set of highly specific 'schemas' based on features common to the child and adult form. For example, in the 'nasal structure' target forms all shared nasality: the infant systematically produced a reduplicated structure with the nasal consonant /ɲ/: *finger* as [ɲẽːɲẽː], *window* as [ɲeːɲeː], and *another* as [ɲaɲa].

The systematic implementation of these prosodic structures leads infants to acquire lexical items that match their limited articulatory and planning capacities. Many early words correspond to the canonical forms produced in babble, and so the early lexicon is largely based on well-rehearsed output patterns. The steady accumulation of new words that already fit these patterns – termed 'selected' words (Vihman, in press,b) – enables construction of the early phonological system. Over time, these patterns are overgeneralized to the production of words that do not automatically match these well-rehearsed prosodic structures, and are thus 'adapted' to fit the overriding patterns in the early output. Vihman (2016) interprets this phenomenon as a mismatch between the infant's 'ambition' to produce increasingly complex word forms and the phonological limitations faced in articulatory control, planning and memory.

Vihman's (2016) analysis of bilingual infants' early prosodic structures shows how phonological systematicity drives infants' acquisition of their first word forms, even when the infant is acquiring two separate languages. Her account demonstrates how the "implicit selection of targets [fits] each child's available production resources" (2016, p.83), implying that word acquisition is led by the infant's phonological capacity. In the present study we extend this analysis to onomatopoeic forms. If Vihman's observations are consistent across acquisition, we would expect that onomatopoeia would be acquired in the same way: not based on any iconic or exceptional status that these forms might possess, but instead in line with the phonological structures that the infant is able to produce.

Here we trace infants' acquisition of onomatopoeia in relation to their wider phonological abilities. Following Vihman's (2016) approach, we begin by considering infant production through an analysis of the early words and their prosodic structures, setting aside onomatopoeia to observe any patterns across non-onomatopoeic words. We then analyse onomatopoeia in relation to these patterns. We expect onomatopoeia (e.g. *quack*, *moo*) to be better-suited to the earliest stages of lexical acquisition than non-onomatopoeic words. We hypothesise that onomatopoeia 1) will match the prosodic structures produced most often in infants' earliest words; 2) will fit infants' prosodic structures more so than non-onomatopoeic forms; and 3) will be produced more accurately than non-onomatopoeic forms. These hypotheses draw on findings from Stoel-Gammon and Cooper (1984) amd Kunnari (2002), who found onomatopoeia to be produced with notable accuracy. We also expand on findings from Laing (2014), tracing the acquisition of onomatopoeia across a larger, cross-linguistic sample. We consider the systematicity of infants' early word production through their use of common phonological structures such as consonant harmony and open CV syllables, and analyse their acquisition of onomatopoeia in relation to this. The early production patterns of sixteen infants acquiring six different languages between them will provide a broad dataset for the analysis of onomatopoeic word production, allowing us to observe how these forms relate to the general trajectory of phonological development, and how they compare with non-onomatopoeic words. We will show that these forms are not marginal in language development, but rather fit within the established framework of early production.

Methodology

Data

Our analysis draws from two types of data: diary data, taken from diary studies of early infant production, and video data of infants interacting with their caregivers (Table 1). We also consider caregiver input, taken from the same video data. The combination of two kinds of data allows consideration of word-by-word acquisition (diary data) in relation to on-line word use (video data). Diary data, while highly subjective, offers an invaluable word-by-word view of the early lexicon in its entirety. Video data, on the other hand, cannot provide a comprehensive overview of lexical acquisition, but its objective account of word production in real-time offers a complimentary perspective to the diary accounts, as well as a valuable source of caregiver input data. Data from Annalena (Elsen, 1991) was previously analysed in Laing (2014).

PLEASE INSERT TABLE 1 ABOUT HERE

Diary data. We selected data that included the infant's first 100 words, phonetically transcribed to allow for prosodic analysis. Two datasets (Laura and Trevor) were sourced from CHILDES (MacWhinney, 2000), four were published datasets (Annalena, Hildegard, M and P), and one was an unpublished dataset made available to the author (Maarja). Six were parent linguists' notations, taken for their own research purposes. Trevor's data was gathered by parents who were not linguists, but participants in a wider project (Compton & Streeter, 1977; Pater, 1997). In this case, the parents were speech pathologists, trained in phonetic transcription by the researchers in order to record their infant's speech. We opted to include different languages in this analysis because this a) allows a wider pool of data from which to select appropriate datasets; and b) takes into account the variability in early production that may come about as a result of language-specific prosodic structures (cf. Vihman & Keren-Portnoy, 2013), while at the same time highlighting any general features in onomatopoeia acquisition. Some accounts provide multiple transcriptions for each word type, offering insight into the variability of early word production, while others include only one or two exemplars for each word. This highlights one issue of working with diary data, particularly across different sources, as it is difficult to avoid methodological discrepancies across datasets. To avoid bias in data selection, this analysis includes all variable productions across the diary data, while the combination of video and diary data control for any potential discrepancies across diary datasets.

We analysed the first 100 word types from each dataset, including all tokens noted for each word. Forms reported in the datasets as words but which correspond more accurately to exclamations or vocal gestures, such as the 'demonstrative interjection' [ʔəʔ] (Hildegard, 0;8, Leopold, 1939) and [aː] to represent 'I'm happy' (M, Deuchar & Quay, 2000, p. 119), were excluded from the analysis. All but two datasets provide details of variability in infant productions, although this is more detailed in some accounts than in others. On average, diary data included 111 word tokens for each infant, with Annalena's data providing the most variability (130 tokens recorded across 100 word types), and Hildegard's the least (no variable word tokens). We include all tokens of the first 100 word types in this analysis. We analysed diary data in ten-word 'sessions', from the first to the 100th word type recorded for each infant. This allows a like-for-like comparison of vocabulary development, despite differences in age: Annalena and Trevor were the first to produce their first word, both at 0;8, while Laura was the last to begin word production, at 1;3.

Video data. Alongside the diary accounts, video data provide a complimentary view of early lexical development and phonological production. We draw on data from CHILDES (MacWhinney, 2000), and include nine infants from two corpora (Table 1). These corpora were collected by a single research team, and offer parallel data collection methods in two different linguistic communities (see Demuth et al., 2006; Demuth & Tremblay, 2008 for full details). They were selected for the present analysis because of the age ranges covered and inclusion of two languages spoken by the author. Recordings began at the onset of word production, and included naturalistic interactions between infants and their caregivers for one-hour sessions at bi-weekly intervals. This provides a dense set of data appropriate for observing developments in early phonological production. Video data were available for all but one infant (Théotime, Lyon corpus), but this did not affect our analysis, which drew exclusively from the transcripts (see below). One infant was excluded from our analysis as he was later diagnosed with Asperger's Syndrome (Providence Corpus, Demuth et al., 2006).

We selected three hour-long recordings from each child's data, representing three points in developmental time. We established Session 1 (S1) as the first recording in which the infant produced at least five different word types. This corresponds roughly to the Diary data, with ten word types in S1. Based on estimates showing that infants produce around 25 words in a given session when they have approximately 50 words in their full lexicon (Kunnari, 2002), we then established S3 as the first session in which the infant produced 50 word types, thereby corresponding to roughly 100 words. These criteria did not suffice for two infants – Violet produced 49 word types at 1;08.22 and 137 types in the following session, while William had already produced 52 word types in S1 (Table 2). To ensure consistency across infants, we also considered word combinations in S3. Following Laing (2015), we ensured that all infants produced at least five different word combinations – defined as those with differing lexo-syntactic structure, so *red doggie* and *want juice* would count as different word combinations, while *red doggie* and *red car* or *red doggie* and *blue doggie* would not. All infants were in the early stages of combining words by S3 (Table 2). Finally, we established S2 by calculating the number of days between S1 and S3, and then selecting the recording with the closest mid-point in age between the two. This demarcated developmental and longitudinal consistency across the analysis. The video data was analysed across the three sessions, mapping progress in phonological and lexical acquisition between the developmental milestones outlined above.

PLEASE INSERT TABLE 2 ABOUT HERE

We extracted data for each session using CLAN (MacWhinney, 2000). For each infant, we recorded each word token produced and its phonetic transcription. Infants produced 133 tokens per session (SD=124; S1: M=52.1, SD=55.3; S2: M=94.3, SD=54.3; S3: M=253.8, SD=136.1); Alex produced the most tokens overall (n=748) and Marie the least (n=226). We also extracted caregiver data for each session, including only nouns (including proper names/character names) and onomatopoeia produced by the mother. Mothers produced 432 word tokens on average (SD=289; S1: M=467.33, SD=345; S2: M=372, SD=254; S3: M=457, SD=289); Lily's mother produced the most words (n=3008) and Marie's the least (n=415).

Analysis

For the three bilingual infants, words that differ in the two languages and were acquired separately (for example, *granny* and *abuela* 'grandma' in M's data) were considered as separate word forms, while words that are phonetically similar across the two languages such that the specific language cannot be determined from the infant's form were considered as one. For example, M produces the form [ba] at 1;3 to represent *button*, but it is unclear whether she has acquired the English token of the word, or the similar Spanish word *botón*. Deuchar and Quay (2000) mark this single form as representing both the English and the Spanish target, and then go on to report two further acquisitions, one of *button* and one of *botón*. Here only the first form is included; all further language-specific differentiation between the two forms (first produced as [bʌʔ] and [bɒn] respectively) reflects wider phonological and lexical changes taking place in the infant's output and are not relevant to this analysis. For the video data, all words produced by the infants in the three sessions and transcribed in CHILDES were recorded.

We began by categorizing the data into three word classes: onomatopoeia, 'conventional' words and 'regular' words. We defined onomatopoeia as words that represent a sound from the environment. This includes animal and engine noises, words that imitate other sounds such as *bang* and *boom*, and *yum yum* to represent an eating sound (Werner & Kaplan, 1963). In some cases we encountered onomatopoeia that were individual to particular infants (e.g. Trevor's *razor* [n:n:]); we included these in the analysis so long as they had a phonological form. Forms without clear prosodic structures (e.g. Annalena's onomatopoeia *pig* is described as 'grunting', Elsen, 1991) could not be considered in our analysis, and were omitted them from the data. Analysis of 'wild' forms (Rhodes, 1994) is beyond the scope of the present study, but will be discussed alongside our results below. We coded any 'conventional' words for comparison with onomatopoeia. In the infant data, a word was considered to be conventional if either a) it had an onomatopoeic equivalent on the CDI form (i.e. in 'sound effects and animal sounds', Fenson et al., 1994) and so we could expect the onomatopoeia to be acquired in early production, or b) if the infant produced a corresponding onomatopoeia as well as the conventional word in their dataset. Conventional words are forms for which we can assume (or have evidence to show) that the infant knows the corresponding onomatopoeia. These were almost exclusively animals and vehicles but also words such as a *clock* (onomatopoeic form *ticktock*): *ticktock* is not on the CDI, but since Naima produces both *clock* and *ticktock* in S3, we include *clock* as a conventional word for her data. Furthermore, if an infant acquired *truck* it would be coded as conventional even if they did not acquire *vroom* because *vroom* is included on the CDI. See supplementary data (SD1) for a full list of onomatopoeic and conventional words in the analysis. All other non-onomatopoeic words were classed as 'regular' words. Regular and conventional words will be combined for the majority of our analysis. The same word classes and criteria for onomatopoeic, conventional and regular word classification also apply to caregiver data.

Prosodic Structures

Our analysis follows Vihman's (2016) approach to the classification of 'prosodic structures'. Prosodic structures represent systematic phonological patterns that occur in individual infants' data. Common prosodic structures include harmony features, such as consonant harmony and reduplication, as well as structural features such as open CV syllables and vocalic onsets (Vihman, 2016). These are thought to derive from the rudimentary vocal gestures produced during babbling (Davis, MacNeilage, & Matyear, 2002; Kent, 1992).

The prosodic structure of each word produced by each infant was annotated by the author (see Table 3). A trained research assistant, blind to the original annotations, re-annotated 12% of the data (n=192); transcribers' agreement was 78% (Cohen's kappa=.7). Disagreements were resolved through discussion.

PLEASE INSERT TABLE 3 ABOUT HERE

Following Vihman (2016), structures accounting for ≥10% of an individual infant's word tokens were termed 'OWN' structures for that infant (onomatopoeia excluded to avoid bias in favour of our hypotheses). Words of the same type that differed in prosodic structure were classed as different tokens. For example, Alex produced 126 instances of *pop* in S2; 113 with the structure CV (/ba/, /bʌ/, /pʌ/) and 13 with consonant harmony (/bab/). To maintain consistency in type and token counts across video and diary data, we considered forms with different structures to be different tokens, while forms with the same structure – regardless of how many times they were produced in a session – were not differentiated. In the above example, Alex would be classed as having two different tokens of the type *pop*, one CV and one CH.

Altogether, we identified nine prosodic structures. Four were common across infants (Figure 1): as reported elsewhere (Kent, 1992; Smith, 2004; Vihman, 1978, 2016), the open CV structure and CH, were most common. All infants had a CH structure, and all but P had a CV structure. Five structures were specific to individual infants (Table 3); these will be classed as one structure, 'IND'. Reduplication was common across the data, and was an OWN structure for six infants (Annalena, Hildegard, Laura, Marie, P, Trevor). However, for most of our analysis we aggregate CH and RED as a general harmony structure (hereafter CH unless otherwise specified). Results were consistent when CH and RED were considered as separate structures.

Vihman (2016, 2017, in press,a,b) refers to 'selected' and 'adapted' words, whereby infants' earliest words are closely matched to the adult form, and may reflect accuracy in production ('selected'). Over time, the pattern force of their OWN structures leads them to acquire words that do not naturally fit that form, and thus are 'adapted' to meet their output preferences. In the present analysis we follow Vihman (2016) in determining the extent to which each word is 'selected' for its producible prosodic structure, or 'adapted' by the infant in order to make it more producible. As Vihman (in press,b, p.46) explains, selected forms are "roughly accurate, within the resources available to the particular child". With this in mind, consonant clusters, word-final fricatives, and adult-like vowels and voicing are not produced reliably in early phonological development, and so forms such as *cheese* [tiːs] (Lily, S3), *bath* [bɑː] (Alex, S1), *hot* [hæt] (Naima, S3) and *dog* [dɑk] (Violet, S3) were all classed as selected. Words were considered to be adapted if the infant form differed in syllable structure (*piggie* [gɪ] – Violet, S3), the presence/absence of a supraglottal onset consonant (*Lego* [ʌgo] – Naima, S2), the absence of a stop consonant in coda position (*duck* [dʌ] – Lily, S3), or adaption in place or manner of articulation (*kitty* [kiki] – Trevor, S3; *Nase* 'nose' [nɑːmə] – Annalena, S6). We considered selection and adaptation for OWN structures only, since a word cannot be selected/adapted for a structure that is not frequent in the output. A trained research assistant re-annotated a 12% sample of the data to compare with the author's original classifications. Agreement was 66% (Cohen's kappa=.35); again, disagreements were resolved through discussion.

We also analysed caregivers' prosodic structures, with some adjustment to the criteria set out above (see SD2). Words with the structure C1VC1(V) were classed as CH if at least two consonants at syllable onset/coda were consistent in place of articulation (for example, *banana*, *crackers*, *doggie*, *toys*). Reduplication was classed as the full repetition of a syllable, allowing voicing changes in the C1 (e.g. *papa*, *backpack*). CV structures had a word-final oral/nasal vowel, or a liquid (e.g. *door*, *toy*, *pain* 'bread'). VCV structures included disyllables with one non-liquid supraglottal closure, with either a vowel, liquid or subglottal closure in word-initial position (*water*, *window*, *hibou* 'owl'). This is a broad interpretation of adult phonology, but it allows us to analyse input from the perspective of infant, rather than adult, production.

Results

Combined video and diary data were not normally distributed (types and tokens, *ps*<.05; Shapiro-Wilkes tests), and so when these are considered together, we report statistical analyses with non-parametric tests on untransformed data, unless otherwise stated. Video data was normally distributed, whereas Diary data was not. Statistical tests on individual Diary/Video data will use parametric tests, with log-transformed Diary data. Non-parametric tests will be used to compare proportional data.

We first tested whether there was any effect of sex, record (Video or Diary), or language on infant word production, including all word classes (onomatopoeia, regular and conventional) in the analysis. ANOVAs revealed no effect for sex or language on either the Video (types and tokens) or Diary data (tokens only; all *ps*>.05). There was a significant difference between number of Types (M_{diary}=100, SD_{diary}=0; M_{video}=64.11, SD_{video}=17.47; $F(1,14)=29.09$, $p<.001$) and Tokens (M_{diary}=111.29, SD_{diary}=12.98; M_{video}=400.22, SD_{video}=163.76; $F(1,14)=70.63$, p<001; log-transformed) produced in Video and Diary data.

Production differed across sessions in the Video (but not Diary) data, so we ran one-way ANOVAs to test for an effect of Session on number of Types and Tokens produced. Session (S1 vs. S2, vs. S3) had a significant effect on both measures (Types: $F(1,25)=149.8$, $p<.001$; Tokens: $F(1,25)=254.5$, $p<.001$), and follow-up paired-samples t-tests showed that infants produced significantly more word types (t(15.98)=-6.16, $p<.001$) and tokens (t(15.94)=-5.84, $p<.001$) in S3 than S2. There was no difference between S1 and S2 (*ps*>.05) for either measure.

Prosodic structures in early production

Next we excluded all onomatopoeia from the data to establish the extent to which OWN structures dominated infants' outputs, independent of onomatopoeia. From this subset (n=1472), 80% of tokens matched infants' OWN structures (SD=.08). Marie and Trevor produced the highest proportion of OWN structures (both 93%), and Violet and Naima the lowest (69% and 68%, respectively). We compared the number of tokens produced by each infant in relation to whether they fit their OWN structures (Fit vs. no Fit) using a paired-samples Wilcoxon Signed-Rank test. Results showed that early production was dominated by OWN structures (Est.Diff=-54, $p<.001$).

Figure 2 shows the proportion of OWN structures in each infant's data by structure type. As expected, CH and CV were the most common structures: all 16 infants used CH predominantly, and the CV structure was also dominant for all infants but P. Proportions were variable: 57% of P's forms fit the CH structure, compared to 15% of Alex's output, while 65% of William's early words fit the CV structure, compared with zero of P's forms and 25% of Annalena's output.

PLEASE INSERT FIGURE 2 ABOUT HERE

A Kruskal-Wallis test comparing the proportion of word tokens that fit each structure (CH vs. CV vs. VCV vs. V vs. IND) revealed a significant effect for Structure ($H(4)=51.18$, $p<.001$; regular and conventional words only). Paired post-hoc comparisons showed that infants produced significantly more forms with CH and CV than with V, VCV or IND (all *ps*$<.01$; Wilcoxon Signed-Rank tests, see Table 4), while proportions of CH and CV did not differ. Forms with VCV fit infants' structures in marginally higher proportions than V ($p=.05$), but there was no difference between VCV and IND or V, or between V and IND.

PLEASE INSERT TABLE 4 HERE

Onomatopoeia production in early acquisition

Next, we analyzed infants' early words for the presence of onomatopoeia, this time excluding regular and conventional words from our analysis (n=173). Each infant acquired 9.1 onomatopoeia types on average (SD=5.3; M_{video}=6.3, M_{diary}=12.3), or 2.23 per session (SD=1.5; M_{video}=2.9, M_{diary}=1.95). Onomatopoeia accounted for 10% of all data across the analysis period, or 9% of each infants' data (word types; SD=.05). There was a decline in onomatopoeia production (word types) across sessions in the Diary data (Figure 3); onomatopoeia accounted for 12% of all word types overall (n=95, SD=.05), but this was concentrated in the earlier sessions. Around 25% of words in S1 and S2 were onomatopoeia, decreasing to 13-16% in S3-S5. By S6, onomatopoeia accounted for <10% of new words. P acquired the highest number of onomatopoeia overall, at 21% of his first 100 words, while Laura had the lowest, at only 4%.

Turning to the Video data, onomatopoeia constituted 10% of the infants' outputs (n=78, SD=.08): 10% of all word tokens in S1 (SD=.1), 13% in S2 (SD=.1), and 7% in S3 (SD=.04; Figure 3). Alex produced the highest proportion of onomatopoeia tokens (19%, n=9) and William the highest number overall (n=20, 13%). An analysis of word types revealed similar trends: onomatopoeia constituted 10% of all word types, 12% in S1, 16% in S2, and 7% in S3. William produced both the highest proportion of onomatopoeia types in his output (14%) and the highest number overall (n=14). As expected, Anais produced the lowest, with zero onomatopoeia types.

PLEASE INSERT FIGURE 3 ABOUT HERE

Prosodic structures and onomatopoeia

Next we re-ran the analysis with the full dataset, incorporating onomatopoeia alongside regular and conventional words. Altogether, 80% of infants' early words fit their OWN structures

(SD=.07; Table 5). Again, we found a significant difference between number of OWN vs. non-OWN structures across word tokens (Est.Diff=-60, p<001, paired Wilcoxon Signed-Rank test). Regular words matched OWN structures in 80% of cases, conventional words in 67%, and onomatopoeia in 71% (Table 5). A Kruskal-Wallis test comparing proportion of tokens that fit infants' OWN structures across the three word classes (regular vs. onomatopoeia vs. conventional) revealed no effect for word class (H(2)=1.86, p=.4); early words matched OWN structures consistently across the data.

PLEASE INSERT TABLE 5 ABOUT HERE

Another Kruskal-Wallis test comparing proportion of words that fit each structure across structure type (CH vs. CV vs. VCV vs. V vs. IND) revealed a significant effect for structure (H(4)=53.8, p<.001); similar to our analysis above, CH and CV were significantly more frequent than all other structures (all ps<.001; Wilcoxon Rank Sum tests). Again, use of CH and CV did not differ. See Table 4.

CH and CV structures were dominant in onomatopoeia production (Figure 2), yet fit these structures no more frequently than regular and conventional forms. It may be that a dominant CH or CV structure leads infants to acquire more onomatopoeia overall. To test this, we ran Spearman's correlations comparing the number of onomatopoeia tokens produced by each infant in relation to their use of CH and CV structures. We removed onomatopoeia so as not to bias our analysis if they have a higher tendency to fit the CH or CV structures. Infants who produced more CH and CV structures acquired more onomatopoeia in their early words (ρ=.59, p=.02). Considering the two structures individually, the correlation was significant for CH structures (ρ=.58, p=.02) but not CV (ρ=.02, p>.05).

Onomatopoeic vs. conventional word production

Next we considered conventional words in our data. It is important to note that these were uncommon in the dataset, accounting for only 5% of all tokens (n=88). On average, infants produced 4.85 conventional word types (Mdn=5, SD=2.7; M_{video}=4.7, SD_{video}=3.2, M_{diary}=5, SD_{diary}=2.2). We tested for trends in onomatopoeia and conventional word production in our data: did infants with more onomatopoeia also produce more conventional words? We ran Spearman's correlations to compare number of onomatopoeia and conventional word types and tokens produced 1) across all data, 2) within Record types, and 3) within sessions (by data type). There was a significant negative correlation between onomatopoeia and conventional word types in the Diary data (ρ=-.91, p<.01), but this did not hold for tokens. No further correlations were found (all *ps*>.05). Paired Wilcoxon Signed-Rank tests revealed a significant difference between the number of onomatopoeia and conventional word types (Est.Diff.=-4.4, p<.01) and tokens (Est.Diff.=-5.5, p<.01) produced by each infant across the data. Infants produced almost twice as many onomatopoeia tokens, and more than double the number of onomatopoeia types than their conventional counterparts. See Figure 4.

PLEASE INSERT FIGURE 4 ABOUT HERE

Table 6 shows all onomatopoeic-conventional word 'pairs' in our data – that is, all conventional words in the data are listed alongside their corresponding onomatopoeia. Owing to a low sample, we include all pairs here, including those that are not OWN structures. Infants acquired the onomatopoeia and its conventional equivalent in only 25 instances; 27% of all conventional words had a matching onomatopoeic counterpart in a given infant's data, compared with only 14% of onomatopoeia.

PLEASE INSERT TABLE 6 ABOUT HERE

Here we turn to Vihman's (2016) consideration of 'selected' and 'adapted' words. Of the 25 pairs, 17 conventional words had an adapted token (shown in bold) or a token that did not fit a structure (NONE), while only five onomatopoeia were adapted or did not fit a structure. Across the data, more than half of word tokens (OWN structures only) were selected (M=.54, Mdn=.53, SD=.08). Alex had the highest proportion of selected words, with 66%, and Trevor the lowest, at 41%.

Considering the data by word class, 80% of onomatopoeia were selected (Mdn=.75, SD=.16), and 50% of conventional words (Mdn=.33, SD=.32). Regular words were selected in 52% of cases (Mdn=.53, SD=.09), though numbers were very low in this class (n=88; see Table 5). Addressing Hypothesis 3, which proposed that onomatopoeia will be produced more accurately than non-onomatopoeic forms, we compared the proportion of selected words across word class (regular vs. onomatopoeic vs. conventional) with a Kruskal-Wallis test, to find a significant effect for word class ($H(2)=14.24$, $p<.001$). Follow-up paired-samples Wilcoxon Signed-Rank tests showed onomatopoeia to be selected in significantly higher proportions than both regular (Est.Diff.=.28, $p<.01$) and conventional words (Est.Diff.=-.44, $p<.01$). Conventional and regular words did not differ (Est.Diff.=-.13, $p>.05$).

Onomatopoeia and conventional words in the input

Finally, we considered the infants' production to that of their caregivers. Total number of word types produced by caregivers and their infants correlated significantly (Spearman's: $\rho=.7$, $p=.03$), but tokens did not ($\rho=.48$, $p=.19$). In contrast, number of onomatopoeia tokens correlated significantly ($\rho=.78$, $p=.01$), whereas types did not ($\rho=.46$, $p=.22$). There was no correlation between conventional word production in infant and caregiver data for either types or tokens ($ps > .1$). Caregivers produced 28 onomatopoeia types on average (SD=8.8) and 19.3 conventional words (SD=6.1), or 126.66 onomatopoeia tokens (SD=74.6) and 67.4 conventional tokens (SD=30.8). Paired-samples Wilcoxon Signed-Rank tests showed that caregiver produced significantly more onomatopoeia than conventional words for both types (Est.Diff.=-11.46, p=.03) and tokens (Est.Diff.=-49, p=.01).

We then analysed caregiver input in relation to their infants' prosodic structures. Across all caregiver data (onomatopoeia included) three structures were present in all infants' inputs: CH/reduplication, VCV and CV. However, while infants' production was dominated by these structures, these three structures accounted for 41% of all caregivers' word tokens (SD=.05); 54% of caregiver input did not fit a structure (SD=.04). Descriptive statistics did not change substantially when onomatopoeia were excluded (No structure: M=.57, SD=.04; common structures: M=.39, SD=.05). CH was the most common of the four structures, accounting for 19% of input data on average (SD=.06; onomatopoeia included). When grouped with reduplication, the two accounted for 22% of input words (SD=.04; RED only: M=.04, SD=.04). Mirroring trends from the infant data, CV was second most common (M=.14, SD=.05), and VCV the least common (M=.04, SD=.04). We ran Spearman's rank correlations comparing infant and caregiver production of each structure type, including all tokens: onomatopoeia were included, as were both OWN- and non-OWN structures to account for all production. Use of CH tokens correlated across caregiver and infant data (ρ=.77, p<.05), but this was marginal for word types (ρ=.63, p=.07). There were no correlations for any of the other structures for either types or tokens (CV, RED, VCV, CH+RED; all ps>.08). The same tests with onomatopoeia removed from the data were all non-significant (CH tokens: ρ=.64, p=.07; all other ps>.08). The common structures in caregivers' data thus did not determine their infants' use of prosodic structures.

Discussion

Infants produce a high number of onomatopoeia in their early words, and previous research suggests that the iconic properties of these forms may support early word learning (Imai & Kita, 2014; Perry et al., 2015). We considered onomatopoeia from a production perspective, to examine whether the acquisition of onomatopoeia could be determined by their simple phonological structures. First replicating findings from Vihman (2016), we found that the early lexicons of the 16 infants in our analysis were dominated by a set of output patterns, or prosodic structures (CH, CV, VCV and V, as well as individual structures). The same output patterns were dominant across infants' onomatopoeia production, and were selected to fit these structures more often than non-onomatopoeic forms. Onomatopoeia offered an automatic fit to the prosodic structures that were most common in the data.

We first observed the kinds of structures that occurred most often in the infants' production. CH and CV were dominant across the data, and, in line with our first hypothesis, onomatopoeia consistently fit these structures, in 71% of cases. Contrasting with expectations drawn in Hypothesis 2, while onomatopoeia matched infants' OWN structures to a high degree, they did not differ from non-onomatopoeic forms in this respect. This speaks to the systematicity of the early vocabulary, and suggests a strong phonological motivation across early lexical acquisition. Supporting our third hypothesis, onomatopoeia were selected significantly more often than both regular and conventional words. This presents further evidence towards a phonological account for onomatopoeia acquisition, suggesting that infants exploit the simple phonological structures inherent in these forms. The onomatopoeia in our data naturally fit the structures that were most common in infants' early words; they did not require adapting and thus were produced with relatively high accuracy. This is consistent with Kunnari's (2002) observations, as Finnish infants' production of onomatopoeia was more accurate than the rest of the early vocabulary.

We might expect that infants who heard more onomatopoeia in their input would produce more onomatopoeia themselves, and to some extent this was the case in our data: mothers who produced more onomatopoeia tokens had infants who produced more onomatopoeia tokens (but not types). This is consistent with previous research, which shows that input is an important determiner of the prosodic structures produced in infants' early words (Vihman & Keren-Portnoy, 2013): Wauquier and Yamaguchi (2013) show that French infants tend to produce open CV syllables in very early production, evolving to a VCV or CH structure over time. This is supported in our data: as Figure 1 shows, all French infants (Anais, Marie, Nathan, Theotime) produce VCV as an OWN structure. Language-specific trends in infants' prosodic structures have also been noted elsewhere: for geminate medial consonants in Lebanese-Arabic (Khattab & Al-Tamimi, 2013), Italian (Vihman & Majorano, 2017) and Finnish (Savinainen-Makkonen, 2013), and consonant clusters in Polish (Szreder, 2013). However, aside from CH, which was used across all infant and caregiver data, there was no correlation between the use of a specific structure in the infants' input and their early words. Combined with the fact that overall onomatopoeia production was correlated between caregiver and infant data, these results point to a combined effect of input features and phonological development in lexical acquisition.

Recent studies suggest that caregivers may draw on iconicity in interactions with their children to provide a more supportive input. Perniss, Lu, Morgan and Vigliocco (2018) show that caregivers exploit iconicity to support the development of referential understanding. They observed child-directed signs produced by deaf mothers to find that signs with the highest iconicity ratings were modified (i.e. rendered more iconic) most frequently, and that this was most frequent when the referent was not present. Perry and colleagues (2018) show a similar effect in that the words produced most frequently in child-directed speech tended to be more iconic. Importantly, this was reflected in their children's speech, and results were consistent when onomatopoeia were removed. Looking at onomatopoeia specifically, Laing, Vihman and Keren-Portnoy (2017) show that they are more salient than their conventional counterparts in speech directed at pre-linguistic infants. Onomatopoeia were produced with a higher pitch and wider pitch range than conventional words, as well as being longer in duration. Furthermore, the presence of reduplication in onomatopoeia meant that infants heard twice as many tokens of a form such as *woof* (i.e. *woof woof*) than of *dog*, despite the fact that overall word counts did not differ. Altogether, findings show that onomatopoeia are more salient and more frequent in the input, as well as matching the structures that infants produce most often in their output. It is perhaps no surprise that they are acquired in such high proportions in early development.

The structures observed in our dataset may have supported infants to acquire early words that are easy to recall, plan and produce, and two structures stood out as being particularly important in early production: CH and CV. Vihman (2016) observes the same trends in her data, stating that "children everywhere are constrained by the same limitations on articulation, speech planning and memory" (p.71). Smith (1973) goes so far as to claim that consonant harmony and reduplication occur as phonological 'universals' in infants' early words, while Ingram (1974) refers to these structures (termed 'assimilated' forms in his account) as common phonological processes in early language production. Infants' early tendency towards the production of CV syllables has also been posited as a 'universal' (Kent, 1992) owing to infants' early physiological predisposition towards mandibular movements, which bring about canonical babbling (see Davis & MacNeilage, 1995). In Vihman's (1992) study of early syllable production across four languages, more than half of infants' first word types consisted of CV syllables that had been practiced in the pre-linguistic phase, highlighting the continuity between pre-linguistic vocalisations and the development of the early lexicon. Our analysis suggests that infants' early forms, and onomatopoeia in particular, match the structures that are ubiquitous in early production, perhaps owing to their babble-like properties. Single open syllables (*moo*, *baa*) and reduplicated structures (*woof woof*, *quack quack*) are particularly dominant in onomatopoeic forms. This is consistent across languages: 'moo' occurs as *moe* in Dutch, *muh* in German and Greek, *mu* in Hungarian and *baeh* in Urdu, and 'quack quack' as *kwak kwak*, *quak quak*, *pa pa*, *háp-háp* and *quak quak* in the same languages, respectively (Abbot, 2004). In Menn and Vihman's (2011) corpus, all 48 onomatopoeia from the ten languages sampled involve consonant harmony/reduplication, a vocalic pattern (V) or a CV syllable in the target form. It seems that their "easily mastered articulatory shape" (Kunnari, 2002, p.133) may make onomatopoeia learnable across languages, and can explain the distinct presence of these forms in early infant data.

However, the systematic use of these structures in our data was not dependent on onomatopoeia; the infants' lexicons were highly systematic when onomatopoeia were excluded, which suggests that systematicity prompted the acquisition of onomatopoeia, rather than the other way around. Results from our correlational analyses provide further support for this, as infants who produced more CH and CV forms in their non-onomatopoeic forms produced more onomatopoeia overall. It seems that having a strong tendency towards CH and CV forms in the output makes an infant more likely to acquire onomatopoeia. As noted by McCune and Vihman (2001), the early forms in our data show clear resemblance to the typical features of babble, and their observed systematicity points to early lexical production that follows the transition from babble to words. This is likely shaped by caregiver feedback to infants' rudimentary forms, which become increasingly word-like and, eventually, lexicalised.

One important factor observed in our data but not considered in depth is reduplication. We collapsed reduplication with consonant harmony for the majority of our analysis; statistical analysis of CH+RED did not differ from analysis of CH alone. Reduplication is dominant in onomatopoeia but often not in wider production, but there may be an important link between the two that supports vocabulary expansion. Returning to Stoel-Gammon and Cooper's (1984) analysis, is it noted that later on in Will's data, around the 50-word point, he begins to systematically implement a consonant harmony pattern in newly-acquired words. The consonant harmony structure may have come about as a result of the strong reduplication pattern in his earlier output. Owing to the similar prosodic features of the two structures, a preference for reduplication in early forms (including onomatopoeia) may enable a move towards the more flexible consonant harmony structure in later production. Smith (1973) considers reduplication to be "a special case of consonant harmony" (p.165), while Ferguson (1983) highlights how this feature may be recruited systematically in some infants' data and less so in others'. In Vihman's (in press,a) analysis of prosodic structures, reduplication was found in all four languages included in the data. However, this did not constitute a strong enough pattern in its own right across French, English and Italian data, so was considered alongside consonant harmony for these infants' analyses. One Finnish infant used reduplication sufficiently to consider this as a prosodic structure, reflecting the presence of reduplicated syllables in Finnish (e.g. *mummu* 'grandma', *paapaa* 'sleeping'), in contrast with the other languages. In our caregiver data, there was some use of reduplication across both English and French recordings, though this was almost exclusive to 'babytalk' words such as *mama* and *booboo* (to refer to an injury).

Onomatopoeia are amongst the few adult words that contain full reduplication in the languages considered in this analysis, though reduplication is common in ideophones across a number of languages (see Dingemanse et al., 2015), used to convey ideas of repetition (for example, *goro* vs. *gorogoro* in Japanese, 'one vs. multiple heavy objects rolling'). Reduplication is a typical feature of ideophones (and thus onomatopoeia), and the fact that onomatopoeia are both iconic and phonologically-suited to infant production might make them especially learnable in early development. We conclude from our analysis that the simple phonological structures of onomatopoeia make them easier to produce, leading to a predominance of these forms in the early lexicon. While this offers an alternative explanation to the iconicity account, we also consider the possibility that the iconic properties of onomatopoeia may drive their simpler phonological properties[2]. As a word class, onomatopoeia represent the repetition of sounds from the environment; that is, sounds that do not originate from the human vocal tract, and thereby do not have linguistic form or articulatory detail. It is unsurprising, in that case, that these imitative forms tend to be simple to produce. Massaro and Perlman (2017) observed early vocabulary

[2] With thanks to the anonymous reviewer for this suggestion.

acquisition in relation to iconicity and ease of articulation (consonants only; quantified along a scale of 1-7). They found that words with higher iconicity ratings were easier to articulate – that is, produced with fewer consonants, and consonants requiring simpler articulation. Furthermore, iconic forms such as onomatopoeia may have a transmission advantage, as iconic forms successfully convey information about sensory experiences (Dingemanse et al., 2015), more so than non-iconic forms (Kanero, Imai, Okuda, Okada & Matsuda, 2014; Laing, 2017). The learnability of onomatopoeia in early development combined with their distinctive communicative function may make them resilient to change (Dingemanse et al., 2015). This would account for their consistently simple phonological structures, as well as the similarities observed across languages.

Another important aspect of onomatopoeia that fell beyond the scope of this study is the use of 'sound effects' such as pitch modulations and articulatory effects (e.g. growling, lip trills), which make these forms sound more realistic. Such sound effects may be particularly important for word learning: in the same way that an infant might be better able to recall the word *bird* when producing a gesture such as waving their hands to imitate wings (Acredolo & Goodwyn, 1988), so too might they recall a word more easily if it is similar to the sound it refers to. Dingemanse and Akita (2017) analysed Japanese ideophones to find that these forms underwent more expressive morphological processes (such as stem repetition and vowel lengthening) and were grammatically less integrated (i.e. more likely to be syntactically optional) than less expressive forms. Extending these findings to onomatopoeia, the lack of grammatical integration of onomatopoeia – especially when paired with sound effects and salient prosodic features – provides them with a range of linguistic cues that sets them apart from the rest of the input. As well as being phonologically accessible, onomatopoeia can stand alone in the speech stream, and function as meaningful utterances in their own right. Furthermore, articulatory precision is not essential: features such as vowel lengthening, voicing alternations and even consonant substitution would not necessarily render an onomatopoeia incomprehensible (see Laing, 2014, for discussion of articulatory precision in onomatopoeia production). We did not analyse infants' use of such forms in our data, but follow-up studies should consider these in relation to the forms in the present dataset, as it is possible that the sets of data would offer different production opportunities in early development.

These results present new evidence towards a role for onomatopoeia in early acquisition, and lead us to propose that the perhaps disproportionate number of onomatopoeia often observed in the early lexicon (Tardif et al., 2008) might be symptomatic of their simple prosodic structures. Rather than being peripheral to the early lexicon, onomatopoeia may serve a functional role in the establishment of a first phonological system – that is, a system that allows for lexical acquisition despite the limitations of the early phonological capacity. Discussion of infants' acquisition of onomatopoeia has largely drawn evidence from perception (e.g. Asano et al., 2015), and word-learning studies have focused on iconicity more generally (e.g. Perry et al., 2015). Here we present an account from production to suggest a multi-faceted profile of onomatopoeia acquisition.

Conclusion

Our results show that the prominence of onomatopoeia in infants' early words is driven at least in part by infants' early production preferences. Rather than setting onomatopoeia apart from the rest of the vocabulary, we show that these forms are acquired in line with the overall trajectory of phonological development, dependent on articulatory capacity at any given point. The fact that onomatopoeia are producible at the very onset of word-use makes them an important feature of overall language development: their producibility allows lexical expansion and production experience, while they simultaneously work within the confines of a minimal phonological inventory.

Acknowledgements

The author would like to thank Marilyn Vihman for provision of the Maarja data and to the anonymous reviewers for their helpful comments on previous versions of this manuscript. This research was funded by the Economic and Social Research Council, grant number ES/J500215/1.

References

Acredolo, L. & Goodwyn, S. W. (1988). Symbolic gesturing in normal infants. *Child Development* **59**, 450-466.

Aoyama, K., & Davis, B. L. (2017). Non-adjacent consonant sequence patterns in English target words during the first-word period. *Journal of Child Language*, **44**, 1065-1087.

Asano, M., Imai, M., Kita, S., Kitajo, K., Okada, H., & Thierry, G. (2015). Sound symbolism scaffolds language development in preverbal infants. *Cortex*, **63**, 196–205.

Behrens, H. (2006). The input–output relationship in first language acquisition. *Language and Cognitive Processes*, **21**, 2–24.

Davis, B. L., & MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech and Hearing Research*, **38**, 1199-1211.

Davis, B. L., MacNeilage, P. F., & Matyear, C. L. (2002). Acquisition of serial complexity in speech production: A comparison of phonetic and phonological approaches to first word production. *Phonetica*, **59**, 75-107.

Demuth, K., Culbertson, J. & Alter, J. (2006). Word-minimality, epenthesis, and coda licensing in the acquisition of English. *Language and Speech* **49**, 137-173.

Demuth, K. & Tremblay, A. (2008). Prosodically-conditioned variability in children's production of French determiners. *Journal of Child Language* **35**, 99-127.

Deuchar, M., & Quay, S. (2001). *Bilingual acquisition: Theoretical implications of a case study*. Oxford: Oxford University Press.

Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Linguistics and Language Compass*, **6**, 654–672.

Dingemanse, M., & Akita, K. (2017). An inverse relation between expressiveness and grammatical integration : on the morphosyntactic typology of ideophones , with special reference to Japanese. *Journal of Linguistics*, **53**, 501-532

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences,* **19**, 603-615.

Elsen, H. (1991). *Erstspracherwerb, Der Erwerb des deutschen Lautsystems*. Wiesbaden: Deutscher Universitaets Verlag.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, i-185.

Ferguson, C. A. (1983). Reduplication in child phonology. *Journal of Child Language*, **10**, 239–243.

Fikkert, P. & Levelt, C. (2008). How does place fall into place? The lexicon and emergent constraints in children's developing phonological grammar. In P. Avery, E. Dresher & K. Rice (eds.) *Contrast in phonology: theory, perception and acquisition* (pp. 231-270). Berlin: Mouton.

Firth, J. R. (1948). Sounds and prosodies. *Transactions of the Philological Society*, 127-152.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **369**, 20130298.

Imai, M., Miyazaki, M., Yeung, H. H., Hidaka, S., Kantartzis, K., Okada, H., & Kita, S. (2015). Sound symbolism facilitates word learning in 14-month-olds. *PLoS ONE*, **10**, e0116494.

Kanero, J., Imai, M., Okuda, J., Okada, H., & Matsuda, T. (2014). How sound symbolism is processed in the brain: a study on Japanese mimetic words. *PLoS One*, **9**, e97905.

Kent, R. D. (1992). The biology of phonological development. In C. A. Ferguson, L. Menn & C. Stoel-Gammon (eds.), *Phonological development: Models, research, implications* (pp. 65-90). Timonium, MD: York Press.

Kern, S. (2010). Les premiers mots du jeune enfant français: analyse quantitative et qualitative du vocabulaire réceptif et productif des deux premières années de vie. *Rééducation orthophonique* **244**, 149-165.

Kern, S. & Gayraud, F. (2010). *L'inventaire français du développement communicatif. Grenoble*, France: La Cigale.

Khattab, G., & Al-Tamimi, J. (2013). Influence of geminate structure on early Arabic templatic patterns. In M. M. Vihman & T. Keren-Portnoy (Eds.), *The Emergence of Phonology: Whole-Word Approaches and Cross-Linguistic Evidence* (pp. 374–414). Cambridge: Cambridge University Press.

Kim, N., & Davis, B. L. (2015). A phonetic approach to consonant repetition in early words. *Infant Behavior and Development*, **40**, 193–203.

Kunnari , S. (2002). Word length in syllables: evidence from early word production in Finnish. *First Language* **22**, 119-135.

Laing, C. E. (2014). A phonological analysis of onomatopoeia in early word production. *First Language*, **34**, 387-405.

Laing, C. E. (2017). A perceptual advantage for onomatopoeia in early word learning: Evidence from eye-tracking. *Journal of Experimental Child Psychology*, **161**, 32-45.

Laing, C. E., Vihman, M., & Keren-Portnoy, T. (2017). How salient are onomatopoeia in the early input? A prosodic analysis of infant-directed speech. *Journal of Child Language*, **44**, 1117-1139.

Leopold, W. F. (1939). *Speech development of a bilingual child, 1: vocabulary growth in the first two years*. Evanston: Northwestern University Press.

MacWhinney, B. (2000). *The CHILDES project: tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Massaro, D. W., & Perlman, M. (2017). Quantifying Iconicity's Contribution during Language Acquisition: Implications for Vocabulary Learning. *Frontiers in Communication*, **2**, 1–14.

McCune, L., & Vihman, M. M. (2001). Early Phonetic and Lexical Development: A Productivity Approach. *Journal of Speech, Language and Hearing Research*, **44**, 670–684.

Menn, L., & Vihman, M. M. (2011). Features in child phonology: Inherent, emergent, or artefacts of analysis? In N. Clements & R. Ridouane (Eds.), *Where Do Phonological Features Come From? Cognitive, Physical and Developmental Bases of Distinctive Speech Categories* (Language F, pp.261–301). Amsterdam: John Benjamins.

Pačesová, J. (1968). *The development of vocabulary in the child*. Brno: Universita J. E. Purkyne.

Pater, J. (1997). Minimal violation and phonological development. *Language Acquisition* **6**, 201-253.

Perniss, P., Lu, J. C., Morgan, G., & Vigliocco, G. (2018). Mapping language to the world: the role of iconicity in the sign language input. *Developmental Science*, **21.**

Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS ONE*, **10**, e0137147.

Perry, L. K., Perlman, M., Winter, B., Massaro, D. W., & Lupyan, G. (2018). Iconicity in the speech of children and adults. *Developmental Science*, **21**, 1–8.

Rhodes, R. (1994). Aural images. In L. Hinton, J. Nichols & J. Ohala, (eds.) *Sound symbolism* (pp.276-292). Cambridge, UK: Cambridge University Press.

Savinainen-Makkonen, T. (2013). Geminate template: A model for first finnish words. In M. M. Vihman & T. Keren-Portnoy (Eds.), *The Emergence of Phonology: Whole-Word Approaches and Cross-Linguistic Evidence* (pp. 362–373). Cambridge: Cambridge University Press.

Smith, N. V. (1973). *The acquisition of phonology: A case study*. Cambridge, UK: Cambridge University Press.

Smith, N. V. (2004). Universal tendencies in the child's acquisition of phonology. In B. Lust & C. Foley (eds.), *First Language Acquisition: The essential readings* (pp.294-306). Oxford, UK: Blackwell Publishing.

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language*, **38**, 1–34.

Stoel-Gammon, C., & Cooper, J. A. (1984). Patterns of early lexical and phonological development. *Journal of Child Language*, **11**, 247–271.

Szreder, M. (2013). The acquisition of consonant clusters in polish: A case study. In M. M. Vihman & T. Keren-Portnoy (Eds.), *The Emergence of Phonology: Whole-Word Approaches and Cross-Linguistic Evidence* (pp. 343–361). Cambridge: Cambridge University Press.

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & Marchman, V. A. (2008). Baby's first 10 words. *Developmental Psychology*, **44**, 929–938.

Vihman, M. M. (1978). Consonant harmony: Its scope and function in child language. In J. H. Greenberg (ed.), *Universals of Human Language* (pp. 281-334). Stanford, CA: Stanford University Press.

Vihman, M. M. (1992). Early syllables and the construction of phonology. In C. A. Ferguson, L. Menn & C. Stoel-Gammon (eds.), *Phonological development: Models, research, implications* (pp. 393-422). Timonium, MD: York Press.

Vihman, M. M. (2014). *Phonological development: The first two years*. (2nd ed.). Oxford: Blackwell.

Vihman, M. (2016). Prosodic structures and templates in bilingual phonological development. *Bilingualism: Language and cognition* **19**, 69-88.

Vihman, M. M. (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology* **108** 1-27.

Vihman, M. M. (in press,a). The development of prosodic structure: A usage-based approach. In P. Prieto & N. Esteve-Gilbert, *Prosodic development in first language acquisition.* TiLAR. Amsterdam: John Benjamins.

Vihman, M. M. (in press,b). *Prosodic Structures and Phonological Templates in Development.* Cambridge: Cambridge University Press.

Vihman, M. M. & Croft, W. (2007). Phonological development: Toward a "radical" template phonology. *Linguistics* **45**, 683-725.

Vihman, M. M., & Keren-Portnoy, T. (Eds, 2013). *The Emergence of Phonology: Whole-Word Approaches and Cross-Linguistic Evidence*. Cambridge: Cambridge University Press.

Vihman, M., & Majorano, M. (2017). The role of geminates in infants' early word production and word-form recognition. *Journal of Child Language*.

Waterson, N. (1971). Child phonology : a prosodic view. *Journal of Linguistics*, **7**, 179–211.

Wauquier, S., & Yamaguchi, N. (2013). Templates in French. In M. M. Vihman & T. Keren-Portnoy (Eds.), *The Emergence of Phonology: Whole-Word Approaches and Cross-Linguistic Evidence* (pp. 317–342). Cambridge: Cambridge University Press.

Werner, H. & Kaplan, B. (1963). *Symbol formation*. New York: John Wiley & Sons.

Winter, B., Perlman, M., Perry, L. K., & Lupyan, G. (2017). Which words are most iconic? *Interaction Studies*, **18**, 430–451.

Table 1: *Data used in the analysis according to Record type (Video or Diary). All video data, as well as Laura and Trevor's diary data, was sourced from CHILDES (MacWhinney, 2000).*

Child	Sex	Language	Source	Record
Annalena*	F	German	Elsen (1991)	Diary
Hildegard	F	German & English (US)	Leopold (1939)	Diary
Laura	F	English (US)	Braunwald (1976)	Diary
Maarja	F	Estonian & English (US)	Vihman & Vihman (2011)	Diary
M	F	Spanish & English (UK)	Deuchar & Quay (2000)	Diary
P	M	Czech	Pačesová (1968)	Diary
Trevor**	M	English (US)	Compton & Streeter (1977), Pater (1997)	Diary
Lily	F	English (US)	Demuth et al., 2006	video
Naima	F	English (US)	Demuth et al., 2006	video
VIOLET	F	English (US)	Demuth et al., 2006	video
Alex	M	English (US)	Demuth et al., 2006	video
William	M	English (US)	Demuth et al., 2006	video
Nathan	M	French	Demuth & Tremblay, 2008	video
Théotime	M	French	Demuth & Tremblay, 2008	video
Anais	F	French	Demuth & Tremblay, 2008	video
Marie	F	French	Demuth & Tremblay, 2008	video

*Data was also considered in Laing (2014) ** Data was collected by parents who were not the researchers.

Table 2: *Infant age in each video session, with number of word types produced in S1-S3, and number of word combinations in S3. Group descriptive statistics (mean and SD) are included.*

Infant	S1	N words	S2	N words	S3	N words	Word combinations
Alex	1;04.27	8	1;07.07	15	1;10.24	52	7
Lily	1;01.02	12	1;05.08	13	1;08.03	72	5
Naima	0;11.26	18	1;01.11	33	1;02.21	60	6
Violet	1;04.20	5	1;07.04	34	1;08.22	49	7
William	1:04.12	52	1;04.27	36	1;06.05	66	5
Théotime	1;0.14	9	1;03.23	51	1;06.23	50	5
Anais	1;02.11	8	1;06.21	13	1;10.23	51	7
Nathan	1;00.24	6	1;07.22	8	2;02.10	50	9
Marie	1;00.02	9	1;02.27	15	1;07.12	56	12
		14.1		**24.2**		**56.2**	**7**
		(14.7)		(14.6)		(8.2)	(2.3)

Table 3: *Prosodic structures classed as OWN in the data, accounting for at least 10% of an infant's output when onomatopoeia were excluded.*

Structure	Description	Examples
Consonant Harmony (CH)	A C1C1(V) structure, with phonological consistency across C1. Voicing or manner may change but place of articulation remains the same. Nasal harmony – whereby the consonants are both nasal but may differ in place of articulation – is also classed as CH.	*egg* /keɪɡ/ - Alex, S1 *that* /dat/ - Laura, S4 *shoes* /ʃuʃ/ - Lily, S3
Reduplication (RED)	Full repetition of a syllable, maintaining all consonant and vowel properties across the two syllables. This structure is aggregated with CH for our analysis unless otherwise specified.	*teddy* /dɛdɛ/ - Annalena, S1 *bubble* /bʌbʌ/ - Alex, S2 *pine cone* /gogo/ - Trevor, S7
CV	An open syllable featuring one supraglottal consonant in word-initial position, proceeded by a vowel or diphthong. Liquids or glottals may occur in word-medial or word-final position.	*bath* /bɑ:/ - Alex, S1 *Opa* 'grandpa' /pa/ - Hildegard, S2 *brioche* /bi:jo/ - Naima, S3
V	A fully vocalic production incorporating either a single vowel (long or short) or a diphthong. Glottals and liquids may also occur, but no supraglottal consonants.	*light* /aɪ/ - Alex, S1 *heiß* 'hot' /ʔaɪ/ - Annalena, S4 *owl* /aʊ:/ - Trevor, S7
VCV	A vowel-initial disyllable. Glottals and liquids may occur in word-initial or word-final position.	*horsie* /ʌsi/ - William, S1 *daddy* /jædi/ - Lily, S2 *ještě* 'once more' /ete/ - P, S4

VC, Violet	A vowel-initial monosyllable. Glottals and liquids may occur in word-initial position.	*allo* ‘hi’ /al/ - Marie, S3 *cake* /ik/ - Lily, S3 *keys* /hi:s/ - Lily, S3
Palatal structure, J (P)	P replaces some laterals with /j/ in early production, later overgeneralizing this pattern to replace other segments.	*balón* ‘ball’ /bɑji/ - S4 *státi* ‘to stand’ /toji:/ - S6 *zpivat* ‘to sing’ /pi:jat/ - S9
CVi structure (Maarja)	Maarja produces a number of open syllables with a final front-rising diphthong. This is described as a ‘front-rising’ template by Vihman (2016), and is shown by Vihman & Vihman (2011) to be proportionate to palatal use in Estonian IDS.	*daddy* /dai:/ - S1 *ball* /bai/ - S2 *(belly)button* /bɤi/ - S3
Lateral structure, L (Annalena)	Words featuring word-final or medial /l/ are particularly dominant in Annalena’s lexicon after the 50-word point. Six of the 10 words acquired in session 6 fit this structure.	*Bild* ‘picture’ /bl̩:/ - S3 *Blume* ‘flowers’ /bal/ - S5 *Schlussesl* ‘key’ /dl̩dl̩ / - S7
Labial-velar structure, W (Violet)	Violet replaces liquids and consonant clusters, as well as single consonants, with /w/ when a form has more than one consonant in the target.	*mommy* /mɑːwi/ - S1 *too big* /duwɪk/ - S2 *book* /wʊk/, /bwʊ/ - S3

Table 4: *Results from Kruskal-Wallis and post-hoc Wilcoxon Rank Sum tests comparing proportion of infants' forms that fit their* OWN *structures across all types (RW+CW+ OW), RW+CW only, and RW, CW and OW individually. Estimated differences shown for each pair of structures compared by Wilcoxon Test.*

Type	**Kruskal-Wallis Test**	Structure	CH	CV	VCV	V
All types	H(4)=53.8***	CH	-			
		CV	-.08	-		
		VCV	.24***	.35***	-	
		V	.34***	.39***	-.04*	-
		IND	.31***	.39***	<-.01	<.01
RW+CW	H(4)=51.18***	CH	-			
		CV	-.13	-		
		VCV	.22**	.33***	-	
		V	.31***	.42***	-.17·	-
		IND	.26***	.39***	-.1	.06
RW[1]	H(4)=51.12***	CH	-			
		CV	-.1	-		
		VCV	.21***	.35***	-	
		V	.31***	.41***	-.05	-
		IND	.28***	.38***	<-.01	<-.01
CW[1]	H(4)=34.88***	CH	-			
		CV	-.4*	-		
		VCV	.22	.5**	-	

		V	.3*	.57***	<-.01	-
		IND	.3*	.57***	<-.01	<-.01
OW[1]	H(4)=48.62***	CH	-			
		CV	.28**	-		
		VCV	.55***	.25***	-	
		V	.56***	.25***	<-.01	-
		IND	.56***	.25***	<-.01	<-.01

·*p*<.06; *p < .05; **p < .01; ***p < .001

OW = onomatopoeia, CW = conventional word, RW = regular word

[1]Bonferroni alpha-error corrections applied: **p*<.017, ***p*<.003, ****p*<.0001

Table 5: *Descriptive statistics (M(SD)) showing proportion of OWN structures according to word class (All, RW+CW, RW, CW, and OW) in Diary and Video data, shown as a single factor and broken down by individual structures.*

Word Class	**N**	**Overall**	**CH**	**CV**	**VCV**	**V**	**IND**
All	1644	.80(.07)	.36(.11)	.42(.16)	.11(.11)	.04(.07)	.07(.11)
RW+CW	1471	.80(.08)	.33(.11)	.44(.17)	.12(.11)	.04(.07)	.07(.12)
RW	1384	.80(.08)	.33(.10)	.43(.17)	.12(.11)	.04(.07)	.08(.12)
CW	88	.67(.37)	.27(.23)	.62(.33)	.08(.15)	.02(.05)	.01(.05)
OW	173	.71(.24)	.57(.25)	.31(.17)	.06(.08)	.03(.1)	.04(.13)

OW = onomatopoeia, CW = conventional word, RW = regular word

Table 6: *OW-CW pairs produced by each infant. Prosodic structures shown for each word form, adapted forms marked in bold. Consonant harmony and reduplication are aggregated as* CH*. Forms that are not* OWN *structures are marked with an asterisk.*

Infant	Target	CW	Structure	OW	Structure
Lily	*doggie*	**dʌdɪ, jæˈjɪ**	**CH**	ʌf, ufi	VCV, VC
	puppy	bʊpi, pʌpi	CH	ʌf, ufi	VCV, VC
Naima	*cat*	**tɛtʰ, kæːkəkə**	**CH**	mau:, mi, mi:jaʊ:, mimi	CV, CH
	kettle	kʌθə	NONE	i:, **mi:**[1]	V, **CV**
Violet	*frog*	**wɑ:k**	**W**	ɹɪbɪtʰ, ɹɪːbəʔ	NONE, VCV
Laura	*doggie*	**gɔgɪ**	**CH**	bauwau, wauwau	CH
Trevor	*kitty*	**kiki**	**CH**	maʊm	CH
	dog	**gʌ, dʌ, dæ**	**CV**	**aʔ**	**V**
Maarja	*duck*	**da**	**CV**	ka, kaka	CV, CH
P	*auto* 'car'	auto	VCV	tidi:	CH
	cici 'pussy'	cici	CH	na:	CV
Hildegard	*duck*	dak	NONE	natnat	CH
M	*car/carro*	ka	CV	b(ɾ)mb(ɾ)m	CH
	cat/gato	**ka**	**CV**	maʊ	CV
	duck	**da**	**CV**	ka, kak	CV, CH
	dog	**da**	**CV**	bəʊwəʊ	CH
	train	**tu**	**CV**	tʃtʃ	CH
	vaca 'cow'	vaʔa, **aʔa**	CV, **VCV**	m:	C*
William	*car*	kɑ, kʌɾə	CV	biʔ, bi:pbi:, bɜbɜ	CV, CH
	doggie	dɑˈgi, **dʌˈti, æˈdi**	NONE, **CH, VCV**	**m̩ˈm̩**	**CH**

	duck	**bʌ**	**CV**	wæ wæ, gʊkwæ	CH
	horsie	ʌsi	VCV	eɪ	V*
	cow	caʊ	CV	mu:	CV
Theotime	*poule* ‘chicken’	pul	NONE	kwẽ	CV
Marie	*chat* ‘cat’	**ga**	**CV**	mja, maːjaː	CV

* Not an OWN structure [1]*squeal*

OW = onomatopoeia, CW = conventional word, RW = regular word